Summary of North Dakota Lake Water Quality Data Compiled for the Nutrient Scientific Technical Exchange Partnership Support (N-STEPS)

Prepared for:

U.S. Environmental Protection Agency
Office of Science and Technology,
Health Ecological Criteria Division
1200 Pennsylvania Avenue, NW
Washington, DC 20460

Prepared by:

Tetra Tech, Inc., Center for Ecological Sciences
1 Park Drive, Suite 200
Research Triangle Park, NC 27709

February 18, 2020

Contents

[TOC \o "1-3" \h \z \u]

Figures

[TOC \h \z \c "Figure"]

Tables

[TOC \h \z \c "Table"]

This memo, part of Task 4 under the NSTEPS North Dakota lakes work plan, provides a brief summary of the available data and a sharing and compilation plan. In addition, we provide a preliminary analysis of data gaps based on this summary.

1. DATA SOURCES

Four Excel files were supplied to NSTEPS. The "AllLakeNutrientData", "AllProfileData', and NLA datasets are referred to as "Grab", "Profile", and "NLA" in subsequent plots. [REF _Ref11772562] summarizes the data files.

Table [SEQ Table * ARABIC]. Summary of data files

File Name	Description of Data	Format	Depths	Years
AllLakeNutrientData.xlsx	Nitrogen, phosphorus, bacteria, and TSS	Long	Varies	1991–2019
AllProfileData.xlsx	DO, pH, conductivity, and water temperature	Wide	Depth Profile	1980–2019
NLA2012_WQData_NSTEPS.xlsx	Chlorophyll a, nutrients, chemistry, and metals	Wide	Integrated Water Column	2012–2013
NLA2017_WQData_NSTEPS.xlsx	Chlorophyll a, nutrients, chemistry, and metals	Wide	Integrated Water Column	2017

2. DATA PREPARATION AND REVIEW

2.1. Data Processing

Data were processed in R (R Core Team 2019). Reported values of zero for typically log-transformed variables were replaced with half of the minimum nonzero value. Near-surface data were used in this report unless otherwise stated. *Near-surface* was defined as depths <= 2 m. The NLA data were collected from a vertically integrated water column and were considered near-surface.

Most of the data received was collected after 1990. Data from October 1, 2009 and forward (most recent 10 water years of data) were used, per North Dakota Department of Water Quality (NDDWQ) assessment methodology. Data with at least two growing season samples per variable in the past 10 years were included. The growing season was defined as April–November.

A Site ID common across the datasets is required to pair data spatially. The "AllLakesNutrientData" and "AllProfileData" both use a 6-digit numeric STORET ID. This ID was selected as the standard site ID. The NLA data used a different site ID format. To standardize the site IDs, all data was loaded into GIS, and NLA sites that were adjacent to a STORET ID were related as such. Non-adjacent NLA sites retained their NLA site ID. NLA site IDs varied across the study periods (2012 and 2017) for the same lake and therefore were standardized to the NLA 2017 format. There were 278 unique sites after data processing.

Lakes were identified using the North Dakota Lake Assessment Unit (AU) when available. Some NLA lakes were not able to be matched to a ND Lake AU. Those lakes were assigned an NHD Common ID (COMID) to uniquely identify them. There were 203 unique lakes after data processing.

2.2. Quality Assurance

The “AllLakeNutrientData” and NLA 2017 files identified non-detect data. The “AllProfileData” and NLA 2012 files did not identify non-detect data. Data from those files were assumed detected and were not adjusted. Non-detect data was imputed with ½ the minimum detection limit if present, else the reported value.

2.3. Extreme Values

Data were visually explored using boxplots to check for extreme values. Values of specific conductivity (63,300 uS/cm), nitrate/nitrite (NO_x, 857 mg/L), and total nitrogen (TN, 863 mg/L) were removed.

2.4. Data Aggregation

The raw data were aggregated (averaged) to the site daily average, lake daily average, lake annual average, and long-term average level ([REF _Ref31746182]). For example, lake annual average means that the data for each lake-year-variable combination was averaged into one observation. Averaging across time and space often helps reduce the noise (variability) in statistical water quality models. Growing season (April–September) versions of the above datasets were also generated.

Table [SEQ Table * ARABIC]. Number of observations in processed datasets

Dataset	All Seasons	Growing Season
Site daily average	16,644	15,655
Lake daily average	14,973	14,105
Lake annual average	6,119	3,871
Lake long-term average	4,061	1,861

3. SPATIAL EXTENT

Monitoring sites were spread across the state, with the majority on or to the northeast of a NW/SE axis across the state; fewer sites were located in the southwest of the state ([REF _Ref12022529]).

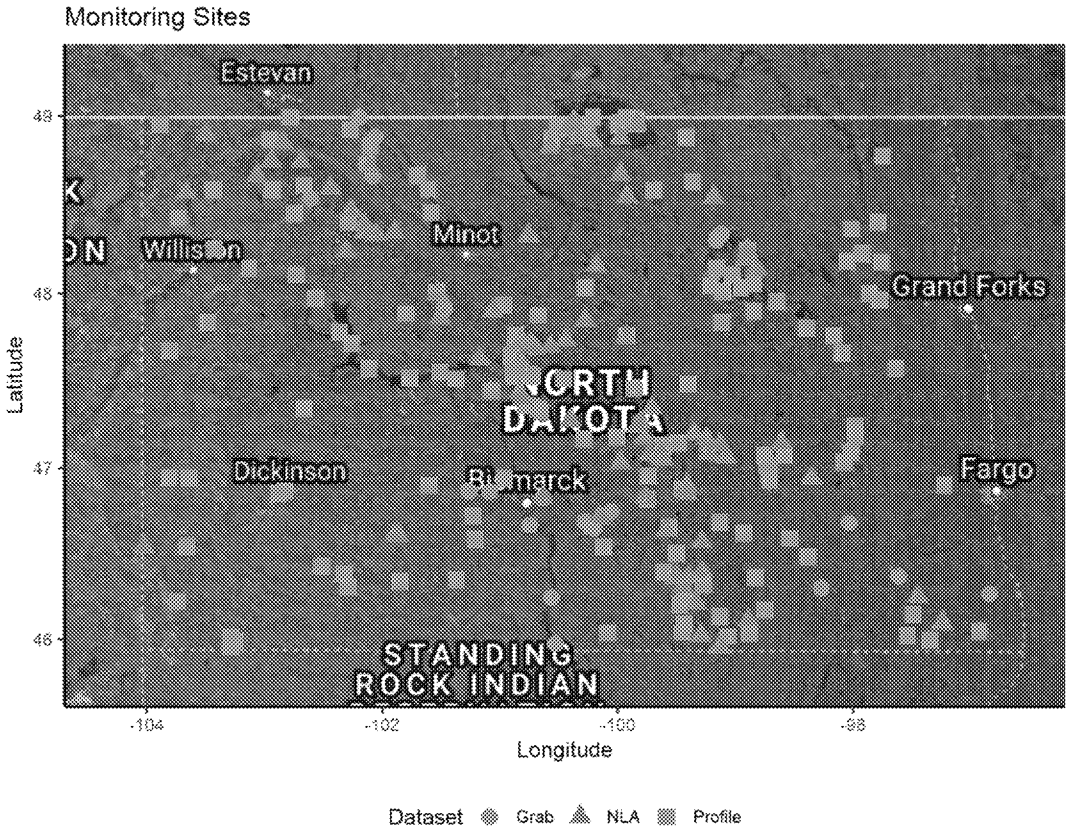


Figure [SEQ Figure * ARABIC]. Map of monitoring sites by dataset

4. TEMPORAL EXTENT

Chlorophyll and Secchi data are temporally sparse compared with dissolved oxygen (DO), nutrients, and total suspended sediment (TSS)([REF _Ref27393118]). Data were most frequently collected between May and October ([REF _Ref27393124]).

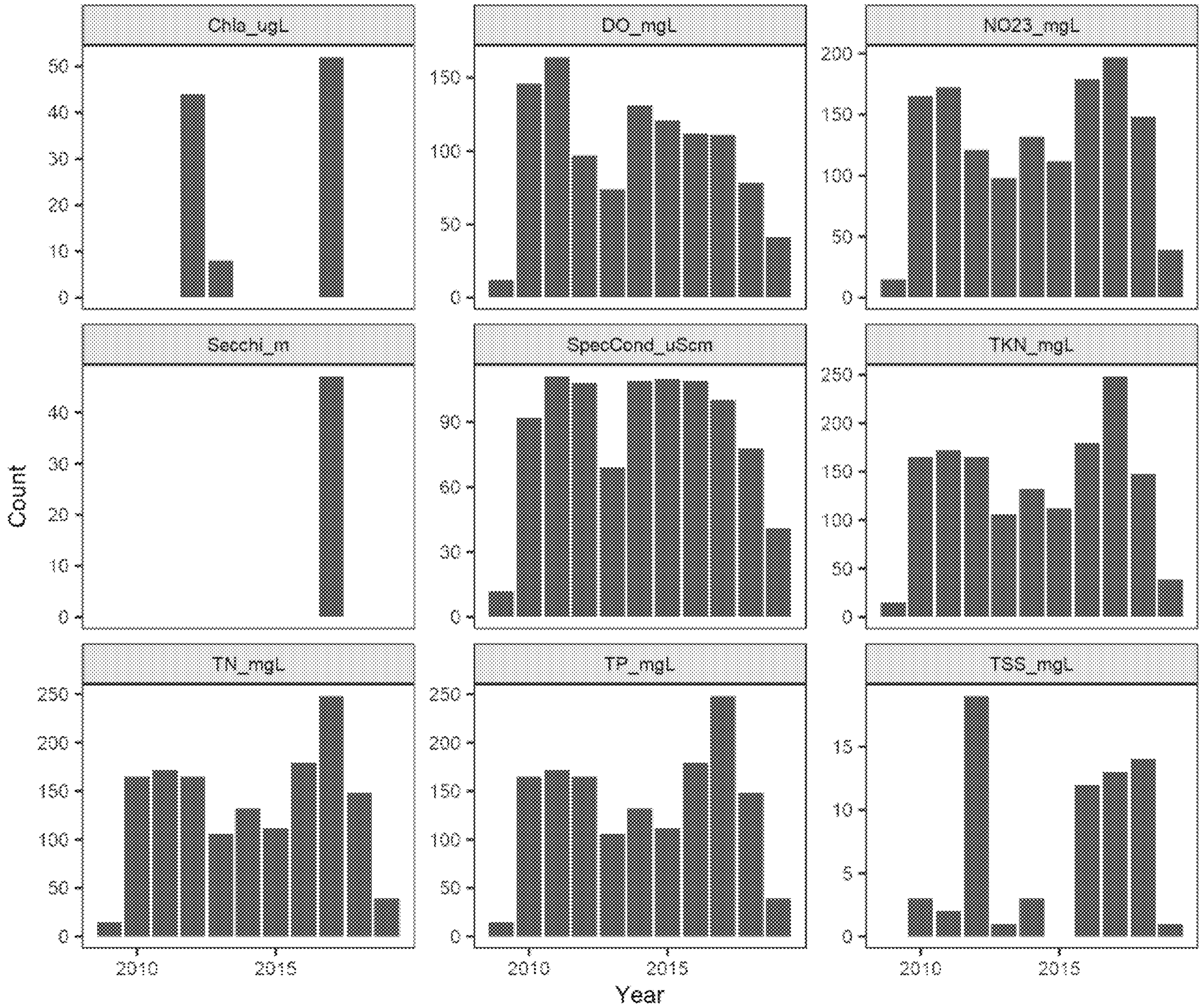


Figure [SEQ Figure * ARABIC]. Growing season lake daily average samples per year.

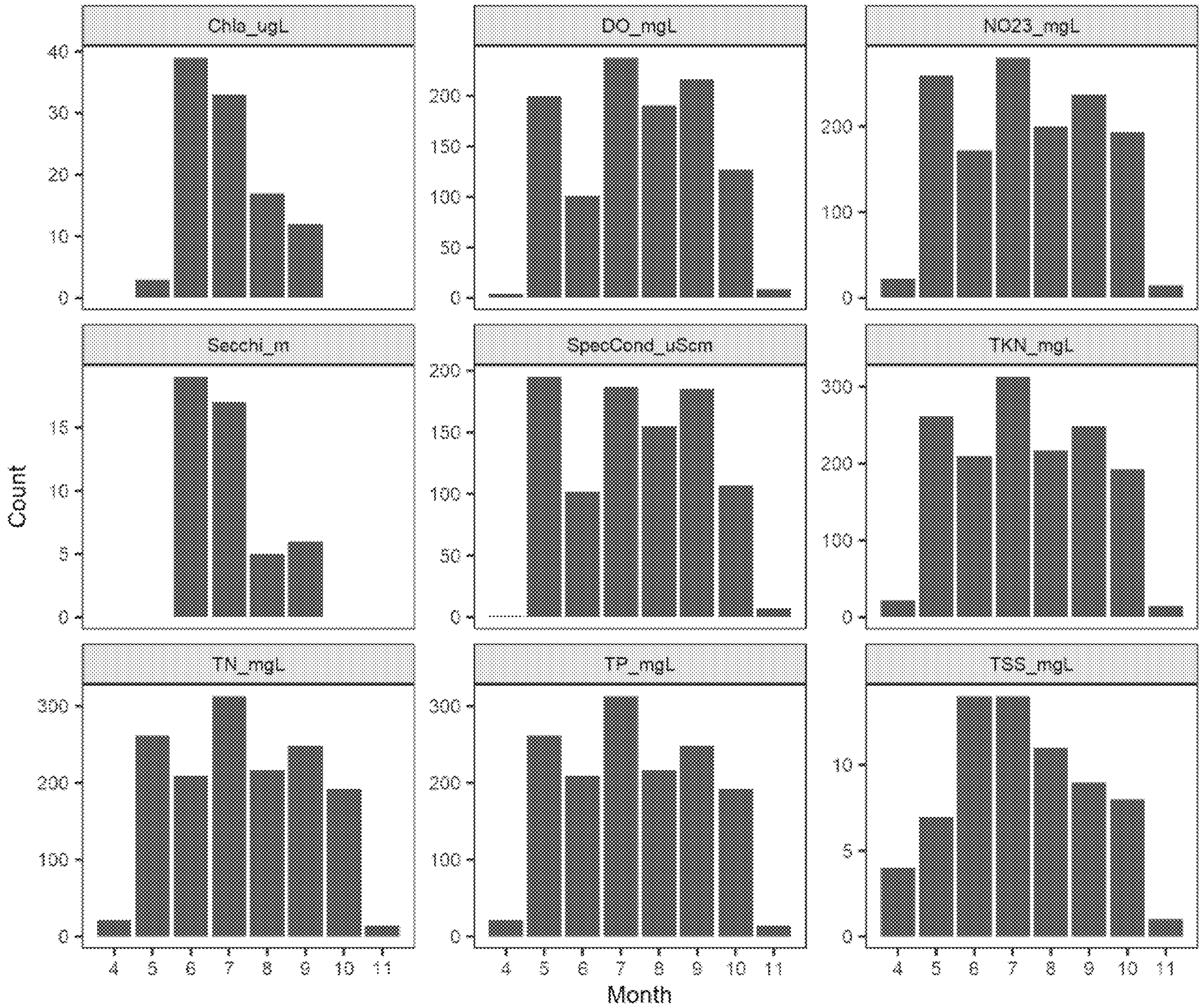


Figure [SEQ Figure * ARABIC]. Growing season lake daily average samples by month.

5. DATA DISTRIBUTIONS

Boxplots were used to explore the univariate distributions of the data. Lake daily averages indicate substantial variability statewide ([REF _Ref27393154]) while long-term averages have more restricted distributions ([REF _Ref31746691 \h]). Monthly patterns appear minimal (in log-space) for many of the variables averaged across all lakes, although there are suspected seasonal trends in DO ([REF _Ref27393171]). The presence of large gradients is hopeful in terms of proposed stressor-response modeling.

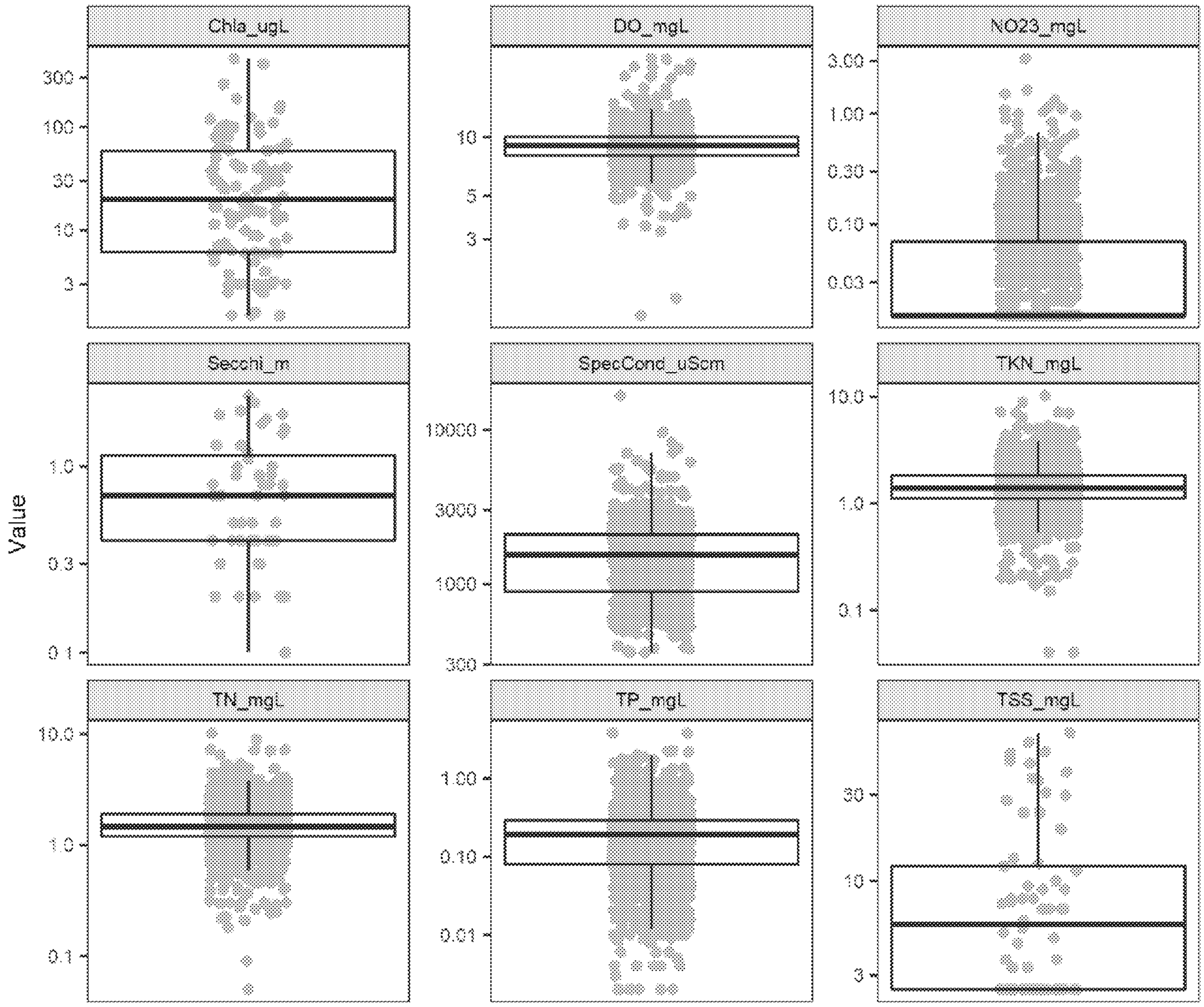


Figure [SEQ Figure * ARABIC]. Boxplots of select variables, using the growing season lake daily average dataset.

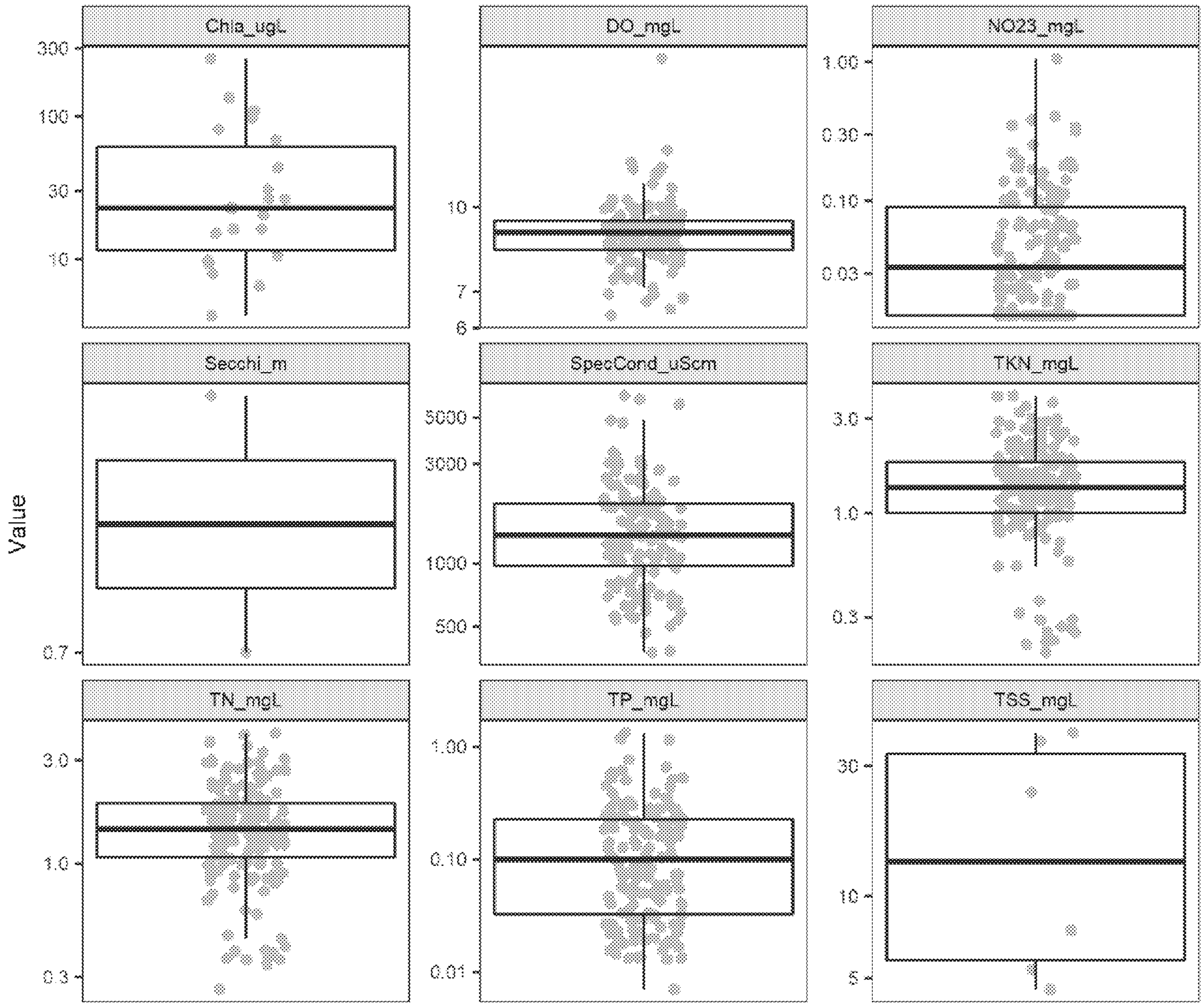
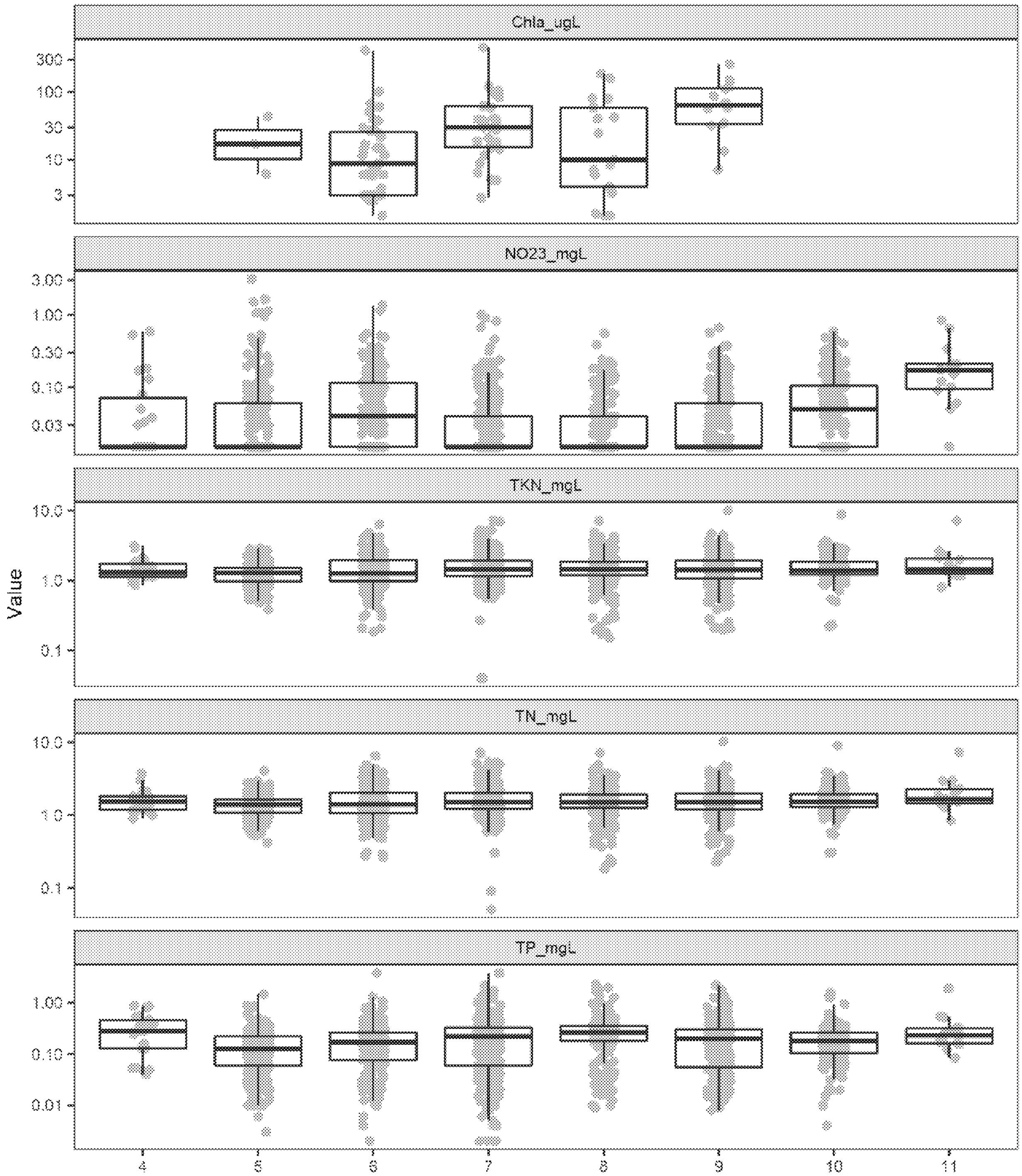


Figure [SEQ Figure * ARABIC]. Boxplots of select variables, using the growing season lake long-term average dataset.


Chla_ugL
300
100
30
10
3
NO23_mgL
3.00
1.00
0.30
0.10
0.03
TKN_mgL
10.0
1.0
0.1
Value
TN_mgL
10.0
1.0
0.1
TP_mgL
1.00
0.10
0.01
4
5
6
7
8
9
10
11

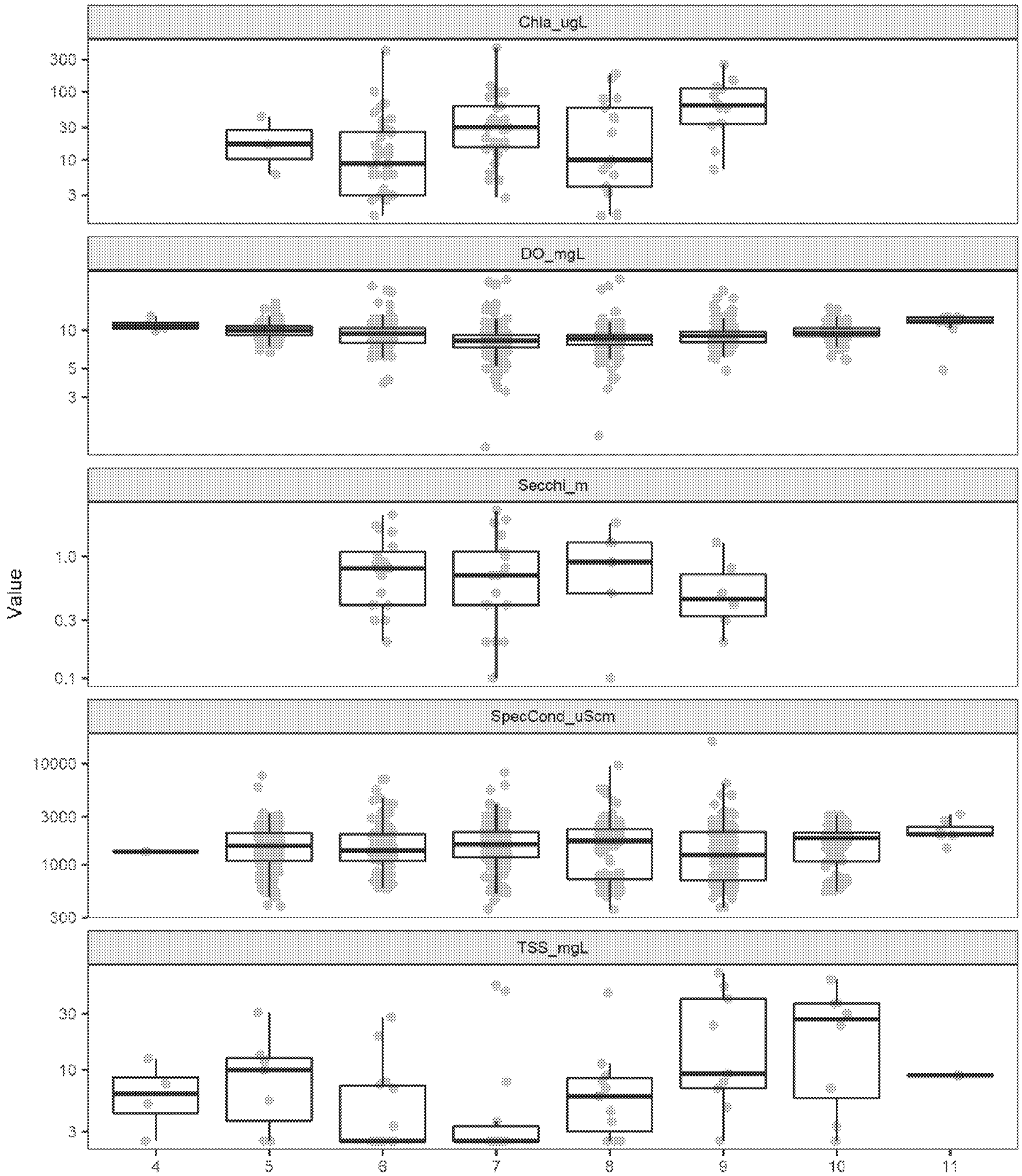


Figure [SEQ Figure * ARABIC]. Boxplots of select variables by month, using the growing season lake daily average dataset.

6. BIVARIATE RELATIONSHIPS

Scatterplots were used to explore bivariate relationships between select variables. Least squares fits were overlaid to visualize trends. Chlorophyll a declines with both DO and Secchi depth and increases with TKN, TN, and TP based on daily average growing season ([REF _Ref27393231], [REF _Ref31746193]). These nutrient responses are the same for long-term average data ([REF _Ref31746586], [REF _Ref31746193]), except for DO and Secchi depth which had insufficient data.The temporally aggregated long-term averages dataset had larger R-squared values than the daily average dataset. TN, TKN, and Secchi predicted chlorophyll better than the other predictors.

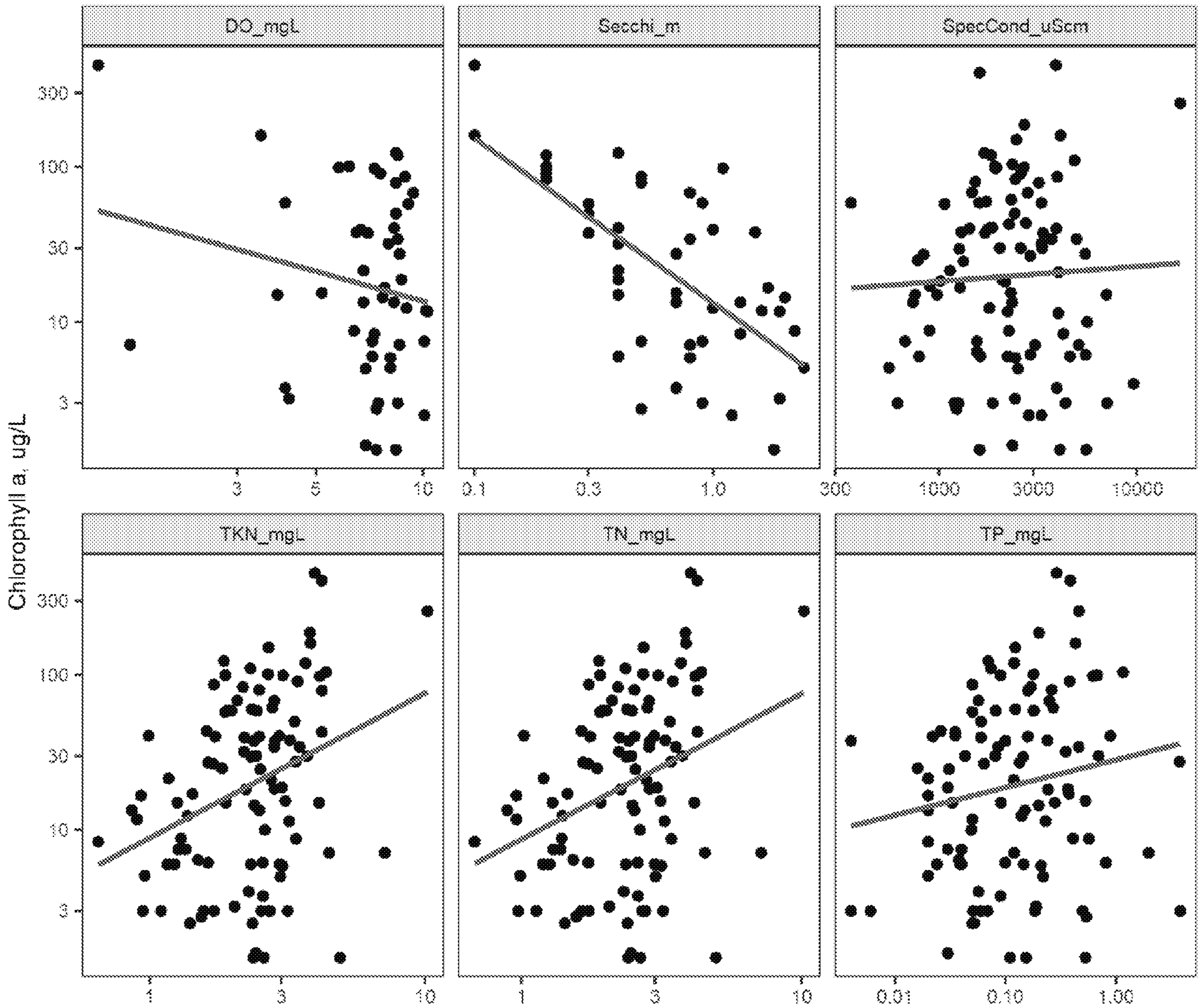


Figure [SEQ Figure * ARABIC]. Scatterplots of select variables with least squares fit (blue), using the growing season daily average dataset

ED_013266A_00023675-00012

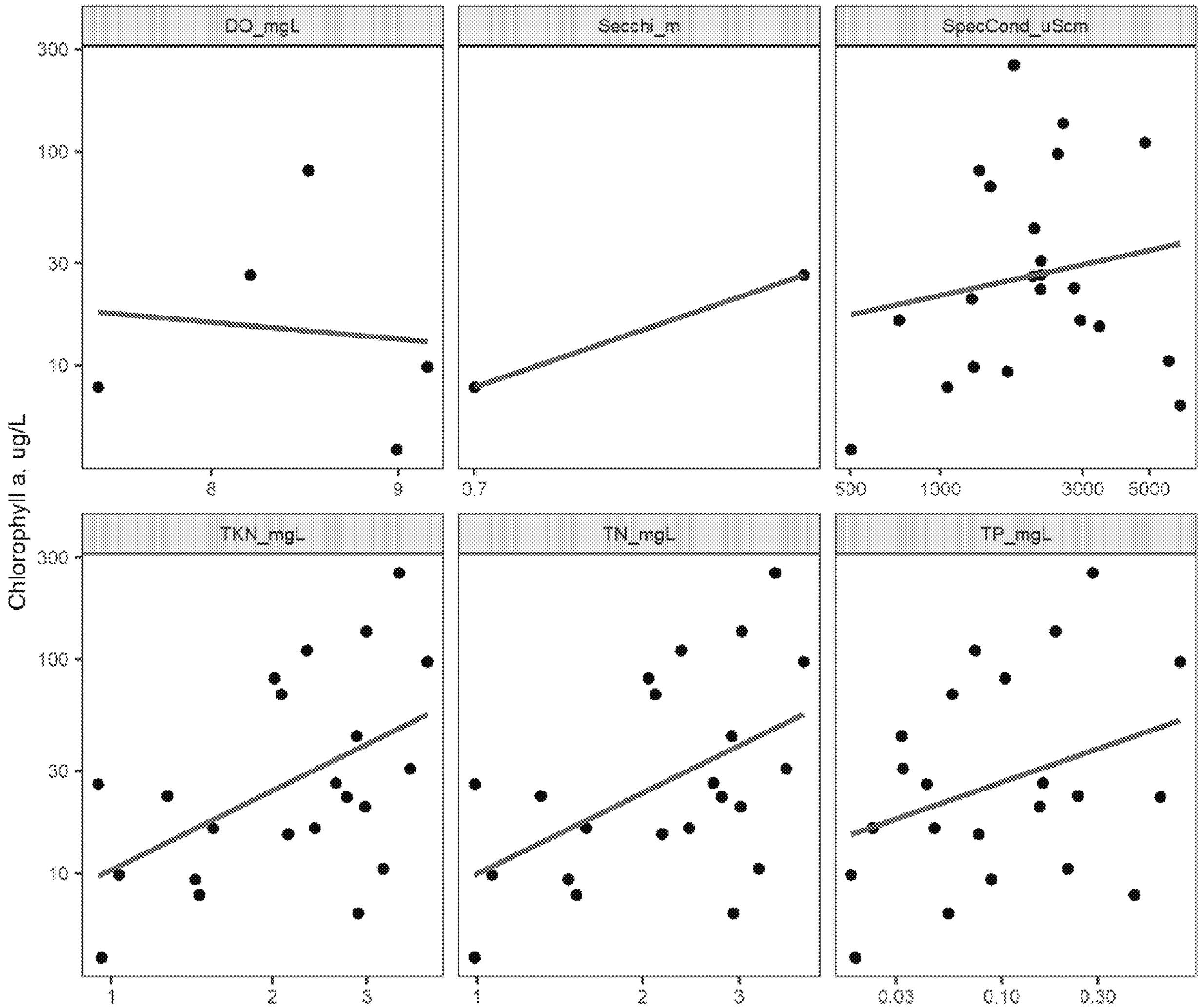


Figure [SEQ Figure * ARABIC]. Scatterplots of select variables with least squares fit (blue), using the growing season long-term average dataset

Table [SEQ Table * ARABIC]. R-squared values for log(Chlorophyll a) vs the following individual predictors

Predictor	Lake-Date, Summer	Lake-Longterm, Summer
log TN	0.10	0.25
log TKN	0.10	0.25
log TP	0.03	0.11
DO	0.02	0.02
log Secchi	0.10	0.25
log Specific Cond	0.00	0.03

7. DATA COMPILATION

NSTEPS proposes to compile the data into "as analyzed" spreadsheets from the above reviewed data. This compilation is not particularly difficult and consists of separate sheets of stand-alone and paired observations (by date and location).

There are some data issues that need to be resolved before the "as analyzed" dataset is considered final. Specifically, no units were provided for the NLA 2012 and 2017 datasets. The units will need to be confirmed.

8. DATA GAPS

In this section, we provide a preliminary evaluation of potential data gaps that could be addressed during the next field season, if still possible. Any recommendation here is being made with only a preliminary run through the data and without having planned or conducted any analyses.

Spatial Gaps

Based on a preliminary look at spatial representation through the coarse mapping we did and without looking at ecoregional breakdowns or breakdowns by lake size, generally the distribution of individual analyte samples across the state is good. It would be helpful to know if the state is concerned about any spatial data gaps. Presumably, the southwestern part of the state is more depauperate in lakes (and water generally), explaining the data densities there.

Temporal Gaps

Based on the preliminary analysis, there is data going back to 1991, but we have focused on the last 10 years. It'd be instructive to know what date range the state is most interested in. By month, the majority of samples are collected between May to September, likely coincident with ice-free seasons. However, enriched dimictic lakes can experience DO sags in winter that can result in alteration of biogeochemical cycles and impacts on aquatic life. It is worth the state considering whether additional winter sampling might be worthwhile to expand on the available data for that time period, especially later in the ice-cover season when sufficient respiration has taken place.

Phytoplankton

We did not have phytoplankton data submitted and just need to confirm that such data do not exist. Many states have well established or are establishing algal sampling programs in lakes for the value of the sensitivity of this assemblage to nutrients, the promise of developing condition indices using algae, and the importance of some algal groups for indicating nuisance or potential nuisance issues. It is worth ND considering the addition of phytoplankton.

Harmful algal toxins

The state may have data on toxins, but these data were not provided. If absent, the state may want to consider adding toxin collection, paired to other water quality data. These toxins have an impact on recreational uses and more is becoming known of their effects on aquatic life. Linking such response measures to nutrients will be important in establishing protective thresholds. Moreover, with the emergence of readily available remotely sensed satellite imagery related to harmful algal bloom events, the ability to sample during potentially toxin producing blooms in lakes already part of the monitoring network, would provide response measures that can be efficiently sampled and linked to preexisting stressor data.

Zooplankton

Zooplankton are not currently being sampled by ND (or the data were not provided), but zooplankton are an important form of aquatic life in lakes and an important link between fish and algae. Shifts in algal assemblage structure with enrichment can affect the palatability of food for zooplankton resulting in shifts in zooplankton assemblages. The state may want to consider adding this assemblage, although this would be considered a lower priority gap to fill.

9. REFERENCES

R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL [HYPERLINK "https://www.R-project.org/"].